

Partial Diffusion Recursive Least-Squares for Distributed Estimation under Noisy Links Condition

Vahid Vahidpour, Amir Rastegarnia, Azam Khalili, and Saeid Sanei, *Senior Member, IEEE*

Abstract—Partial diffusion-based recursive least squares (PDRLS) is an effective method for reducing computational load and power consumption in adaptive network implementation. In this method, each node shares a part of its intermediate estimate vector with its neighbors at each iteration. PDRLS algorithm reduces the internode communications relative to the full-diffusion RLS algorithm. This selection of estimate entries becomes more appealing when the information fuse over noisy links. In this paper, we study the steady-state performance of PDRLS algorithm in presence of noisy links and investigate its convergence in both mean and mean-square senses. We also derive a theoretical expression for its steady-state mean-square deviation (MSD). The simulation results illustrate that the stability conditions for PDRLS under noisy links are not sufficient to guarantee its convergence. Strictly speaking, considering non-ideal links condition adds a new complexity to the estimation problem for which the PDRLS algorithm becomes unstable and do not converge for any value of the forgetting factor.

Index Terms—Adaptive networks, diffusion adaptation, distributed estimation, energy conservation, recursive least-square, partial diffusion, noisy links.

I. INTRODUCTION

WE study the problem of distributed estimation over adaptive networks, in which a set of agents are interacted with each other to solve distributed estimation and inference problems in a collaborative manner. There exist several useful techniques for solving such optimization problems in distributed manner that enable adaptation and learning in real-time. They include incremental [1], [2], [3], consensus [4], [5], [6], and diffusion [7], [8], [9], [10], [11], [12], [13], [14], [15] strategies. The diffusion strategies are effective methods for performing distributed estimation over adaptive networks. In the original diffusion strategy, all agents in the network generate their individual intermediate estimates using the data accessible to them locally. Then, the nodes exchange their intermediate estimates to all their immediate neighbors. However, the most expensive part of realizing a cooperative task over a wireless ad hoc network is usually the data communications through radio links. Therefore, it is of practical importance to reduce the amount of internode communications in diffusion strategies while maintaining the benefits of cooperation.

To aim this, various techniques have been proposed, such as choosing a subset of the nodes [16], [17], [18], [19], selecting

a subset of the entries of the estimates [20], [21], and reducing the dimension of the estimates [22], [23], [24]. Among these methods, we focus on the second method in which a subset of the entries are selected in communications.

In all mentioned works, the data are assumed to be exchanged among neighboring nodes without distortion. However, due to the link noise, this assumption may not be true in practice. Some useful results out of studying the effect of link noise during the exchange of weight estimates, already appear for traditional diffusion algorithm [25], [26], [27], [28], for incremental case [29], [30], [31], and for consensus-based algorithms [32], [33].

In partial-diffusion strategies proposed in [20], [21], the links between nodes are assumed to be ideal. However, the performance of these strategies, due to replacing the unavailable entries by their corresponding ones in each nodes own intermediate estimate vector, are strongly affected under noisy information exchange.

The main contributions of this paper can be summarized as follows:

- (i) Focusing on [21], we consider the fact that the weight estimates exchanged among the nodes can be subject to quantization errors and additive noise over communication links. We also consider two different schemes [20] for selecting the weight vector entries for transmission at each iteration. We allow for noisy exchange just during the two combination steps. It should be noted that since our objective is to minimize the internode communication, the nodes only exchange their intermediate estimates with their neighbors;
- (ii) Using the energy conservation argument [34] we analyze the stability of algorithms in mean and mean square sense under certain statistical conditions;
- (iii) Stability conditions for PDRLS are derived under noisy links, and it is demonstrated that the steady-state performance of the new algorithm is not as good as that of the PDRLS algorithm with ideal links;
- (iv) We derive a variance relation which contains an additional terms in comparison to original PDRLS algorithm that represent the effects of noisy links. We evaluate these moments and derived closed-form expressions for mean-square deviation (MSD) to explain the steady-state performance.
- (v) It is demonstrated through different examples that the stability conditions for PDRLS under noisy links are not sufficient to guarantee its convergence. Considering noisy links adds a new complexity to estimation problem for which the PDRLS algorithm suffers a significant

V. Vahidpour, A. Rastegarnia, and A. Khalili and are with the Department of Electrical Engineering, Malayer University, Malayer 65719-95863, Iran (email: v.vahidpour@gmail.com; rastegar@tabrizu.ac.ir; a.khalili@tabrizu.ac.ir).

S. Sanei is with the Department of Computer Science, University of Surrey, Surrey GU2 7XH, UK (email: s.sanei@surrey.ac.uk).

degradation in steady-state performance for any value of forgetting factor.

The remainder of this paper is organized as follows: In Section II, we formulate the PDLMS under imperfect information exchange. The performance analyses are examined in Section III, where the methods of entry selection matrix are also examined. We provide simulation results in Section IV and draw the conclusions in Section V.

A. Notation

We use the lowercase boldface letters to denote vectors, uppercase boldface letter for matrices, and lowercase plain letters for scalar variables. We also use $(\cdot)^*$ to denote conjugate transposition, $\text{Tr}(\cdot)$ for the trace of its matrix argument, \otimes for the Kronecker product, and $\text{vec}(\cdot)$ for a vector formed by stacking the columns of its matrix argument. We further use $\text{diag}\{\dots\}$ to denote a (block) diagonal matrix formed from its argument, and $\text{col}\{\dots\}$ to denote a column vector formed by stacking its arguments on top of each other. All vectors in our treatment are column vectors, with the exception of regression vectors, $\mathbf{u}_{k,i}$.

II. ALGORITHM DESCRIPTION

A. Diffusion Recursive Least-Squares Algorithm with Noisy Links

To begin with, consider a set of N nodes, spatially distributed over some region, aim to identify an unknown parameter vector, $w^o \in C^{M \times 1}$, in a collective manner. At every time instant i , node k collects a measurement $d_k(i)$ that is assumed to be related to an unknown vector by

$$d_k(i) = \mathbf{u}_{k,i} w^o + v_k(i) \quad (1)$$

where $\mathbf{u}_k(i)$ is a row vector of length M (the regressor of node k at time i) and $v_k(i)$ represent additive noise with zero mean and variance $\sigma_{v,k}^2$ and the unknown vector w^o denotes the parameter of interest.

Collecting the measurements and noise samples of node k up to time i into vector quantities as follows:

$$\mathbf{y}_{k,i} = \text{col}\{d_k(i), \dots, d_k(0)\} \quad (2)$$

$$\mathbf{H}_{k,i} = \text{col}\{\mathbf{u}_k(i), \dots, \mathbf{u}_k(0)\} \quad (3)$$

$$\mathbf{v}_{k,i} = \text{col}\{v_k(i), \dots, v_k(0)\} \quad (4)$$

the objective is to estimate w^o by solving the following weighted least-squares problem

$$\min_{\boldsymbol{\psi}} \|\mathbf{y}_{k,i} - \mathbf{H}_{k,i} \boldsymbol{\psi}\|_{\Lambda_i}^2 \quad (5)$$

The solution $\boldsymbol{\psi}_{k,i}$ is given by [35]

$$\boldsymbol{\psi}_{k,i} = (\mathbf{H}_{k,i}^* \Lambda_i \mathbf{H}_{k,i})^{-1} (\mathbf{H}_{k,i}^* \Lambda_i \mathbf{y}_{k,i}) \quad (6)$$

where $\Lambda_i \geq 0$ denotes a Hermitian weighting matrix. Common choice for Λ_i is

$$\Lambda_i = \text{diag}\{1, \lambda, \dots, \lambda^i\} \quad (7)$$

where $0 \ll \lambda \leq 1$ is a forgetting factor whose value is generally very close to one.

It can be verified from the properties of recursive least-squares solutions [35], [34] that

$$\mathbf{H}_{k,i}^* \Lambda_i \mathbf{H}_{k,i} = \lambda \mathbf{H}_{k,i-1}^* \Lambda_{i-1} \mathbf{H}_{k,i-1} + \mathbf{u}_{k,i}^* \mathbf{u}_{k,i} \quad (8)$$

$$\mathbf{H}_{k,i}^* \Lambda_i \mathbf{y}_{k,i} = \lambda \mathbf{H}_{k,i-1}^* \Lambda_{i-1} \mathbf{y}_{k,i-1} + \mathbf{u}_{k,i}^* d_k(i) \quad (9)$$

Let $\mathbf{P}_{k,i} = (\mathbf{H}_{k,i}^* \Lambda_i \mathbf{H}_{k,i})^{-1}$. Then applying the so-called matrix inversion formula [36][36] to (8) the following recursions for calculating $\boldsymbol{\psi}_{k,i}$ are obtained

$$\mathbf{P}_{k,i} = \lambda^{-1} \left(\mathbf{P}_{k,i-1} - \frac{\lambda^{-1} \mathbf{P}_{k,i-1} \mathbf{u}_{k,i}^* \mathbf{u}_{k,i} \mathbf{P}_{k,i-1}}{1 + \lambda^{-1} \mathbf{u}_{k,i}^* \mathbf{P}_{k,i-1} \mathbf{u}_{k,i}} \right) \quad (10)$$

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{\psi}_{k,i-1} + \mathbf{P}_{k,i} \mathbf{u}_{k,i}^* (d_k(i) - \mathbf{u}_{k,i} \boldsymbol{\psi}_{k,i-1}) \quad (11)$$

Since $\mathbf{w}_{k,i}$ is a better estimate compared with $\boldsymbol{\psi}_{k,i}$, it is beneficial to replace $\boldsymbol{\psi}_{k,i-1}$ with $\mathbf{w}_{k,i-1}$ in (11)

$$\boldsymbol{\psi}_{k,i} = \mathbf{w}_{k,i-1} + \mathbf{P}_{k,i} \mathbf{u}_{k,i}^* (d_k(i) - \mathbf{u}_{k,i} \mathbf{w}_{k,i-1}) \quad (12)$$

Hence, the local estimates are diffused outside of each node's own neighborhood. Then, for every time instant i , each node k performs an adaptation step followed by a combination step as follows:

- 1) *Adaptation*: Each node computes an intermediate estimate of w^o by (10) and (12). The resulting pre-estimates are named $\boldsymbol{\psi}_{k,i}$ as in (13).
- 2) *Combination*: The nodes exchange their local pre-estimates with their neighbors and perform a weighted average as in (14) to obtain the estimate $\mathbf{w}_{k,i}$ (via so-called spatial update).

$$\mathbf{w}_{k,i} = \sum_{l=1}^N a_{lk} \boldsymbol{\psi}_{l,i} \quad (13)$$

The scalar a_{lk} is non-negative real coefficient corresponding to the (l, k) entries of $N \times N$ combination matrix $A = \{a_{lk}\}$. These coefficients are zero whenever node $l \notin \mathcal{N}_k$, where \mathcal{N}_k denotes the neighborhood of node k . This matrix is assumed to satisfy the condition:

$$A^T \mathbb{1}_N = \mathbb{1}_N \quad (14)$$

where the notation $\mathbb{1}$ denotes an $N \times 1$ column vector with all its entries equal to one.

The above algorithm only uses the local input-output data, observed by each node, in the adaptation phase. However, in [15], a more general algorithm has been proposed in which each node shares its input-output data as well as its intermediate estimate with its neighbors and updates its intermediate estimate using all available data. This is carried out via a convex combination of the update terms induced by each input-output data pair. For the obvious reason of minimizing the communications, here, we will only consider the above-mentioned diffusion RLS algorithm.

We model the noisy data received by node k from its neighbor l as follows:

$$\boldsymbol{\psi}_{lk,i} = \boldsymbol{\psi}_{l,i} + \mathbf{v}_{lk,i}^{(\boldsymbol{\psi})} \quad (15)$$

where $\mathbf{v}_{lk,i}^{(\boldsymbol{\psi})}$ ($M \times 1$) denotes vector noise signal. It is temporally white and spatially independent random process with

zero mean and covariance given by $\mathbf{R}_{v,lk}^{(\psi)}$. The quantity $\mathbf{R}_{v,lk}^{(\psi)}$ are all zero if $l \notin \mathcal{N}_k$ or when $l = k$. it should be noted that the subscript lk indicates that l is the source and k is the sink, i.e. the flow of information is from l to k .

Using the perturbed estimate (15), the combination step in adaptive strategy becomes

$$\mathbf{w}_{k,i} = \sum_{l \in \mathcal{N}_k} a_{lk} \psi_{lk,i} \quad (16)$$

B. Partial-Diffusion RLS Algorithm under Noisy Information Exchange

In order to reduce the amount of communication required among the nodes we utilize partial-diffusion strategy proposed in [20], to transmit L out of M entries of the intermediate estimates at each time instant where the integer L is fixed and pre-specified. The selection of the to-be-transmitted entries at node k and time instant i can be characterized by an $M \times M$ diagonal entry-selection matrix, denoted by $\mathbf{K}_{k,i}$, that has L ones and $M - L$ zeros on its diagonal. The proposition of ones specifies the selected entries. Multiplication of an intermediate estimate vector by this matrix replaces its non-selected entries with zero.

Rewriting (16) as

$$\mathbf{w}_{k,i} = a_{kk} \psi_{k,i} + \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{lk} [\mathbf{K}_{l,i} \psi_{lk,i} + (\mathbf{I}_M - \mathbf{K}_{l,i}) \psi_{lk,i}] \quad (17)$$

Each node requires the knowledge of all entries of its neighbors intermediate estimate vectors for combination. However, when the intermediate estimates are partially transmitted ($0 < L < M$), the nodes have no access to the non-communicated entries. To resolve this ambiguity, we let the nodes use the entries of their own intermediate estimates in lieu of ones from the neighbors that have not been communicated, i.e., at node k , substitute

$$(\mathbf{I}_M - \mathbf{K}_{l,i}) \psi_{lk,i}, \forall l \in \mathcal{N}_k \setminus \{k\} \quad (18)$$

for

$$(\mathbf{I}_M - \mathbf{K}_{l,i}) \psi_{lk,i}, \forall l \in \mathcal{N}_k \setminus \{k\} \quad (19)$$

Based on this approach, we formulate a partial-diffusion recursive least-squares (PDRLS) algorithm under imperfect information exchange using (10) and (12) for adaptation and the following equation for combination:

$$\begin{aligned} \mathbf{w}_{k,i} = & a_{kk} \psi_{k,i} + \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{lk} [\mathbf{K}_{l,i} \psi_{lk,i} + (\mathbf{I}_M - \mathbf{K}_{l,i}) \psi_{lk,i}] \\ & + \mathbf{v}_{k,i}^{(\psi)} \end{aligned}$$

where $\mathbf{v}_{k,i}^{(\psi)}$ denotes the aggregate $M \times 1$ zero mean noise signal and is introduced as follows:

$$\mathbf{v}_{k,i}^{(\psi)} = \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{lk} \mathbf{K}_{l,i} \mathbf{v}_{lk,i}^{(\psi)} \quad (21)$$

This noise represents the aggregate effect on node k of all selected exchange noises from the neighbors of node k while exchanging the estimates $\{\psi_{lk,i}\}$ during the combination step.

C. Entry Selection Methods

In order to select L out of M entries of the intermediate estimates of each node at each iteration, the processes we utilized are analogous to the selection processes in stochastic and sequential *partial-update* schemes [37], [38], [39], [40]. In other words, we use the same schemes as introduced in [20]. Here, we just review these methods named *sequential partial-diffusion* and *stochastic partial-diffusion*.

In sequential partial-diffusion the entry selection matrices, $\mathbf{K}_{k,i}$, are diagonal matrices:

$$\mathbf{K}_{k,i} = \begin{bmatrix} \kappa_{1,i} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \kappa_{M,i} \end{bmatrix}, \kappa_{\ell,i} = \begin{cases} 1 & \text{if } \ell \in \mathcal{J}_{(i \bmod \bar{B})+1} \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

with $\bar{B} = \lceil M/L \rceil$. The number of selection entries at each iteration is limited by L . The coefficient subsets \mathcal{J}_i are not unique as long as they obey the following requirements [37]:

- 1) Cardinality of \mathcal{J}_i is between 1 and L ;
- 2) $\bigcup_{\kappa=1}^{\bar{B}} \mathcal{S} = \mathcal{S}$ where $\mathcal{S} = \{1, 2, \dots, M\}$;
- 3) $\mathcal{J}_{\kappa} \cap \mathcal{J}_{\eta} = \emptyset, \forall \kappa, \eta \in \{1, \dots, \bar{B}\}$ and $\kappa \neq \eta$.

The description of the entry selection matrices, $\mathbf{K}_{k,i}$, in stochastic partial-diffusion is similar to that of sequential one. The only difference is as follows. At a given iteration, i , of the sequential case one of the sets $\mathcal{J}_{\kappa}, \kappa = \{1, \dots, \bar{B}\}$ is chosen in advance, whereas for stochastic case, one of the sets \mathcal{J}_{κ} is sampled at random from $\{\mathcal{J}_1, \mathcal{J}_2, \dots, \mathcal{J}_{\bar{B}}\}$. One might ask why these methods are considered to organize these selection matrices. To answer this question, it is worth mentioning that the nodes need to know which entries of their neighbors intermediate estimates have been transmitted at each iteration. These schemes bypass the need for addressing.

Remark. The probability of transmission for all the entries at each node is equal and expressed as

$$\rho = L/M \quad (23)$$

Moreover, the entry selection matrices, $\mathbf{K}_{k,i}$, do not depend on any data/parameter other than L and M .

III. PERFORMANCE ANALYSIS

In this section, we analyze the performance of the algorithm (10), (12) and (20) and show that it is asymptotically unbiased in the mean and converges in the mean-square error sense under some simplifying assumptions. We also provide an expression for mean-square deviation (MSD). We presume that the input data is stochastic in nature and utilize the energy conservation arguments previously applied to LMS-type adaptive distributed algorithms in, e.g., [9], [12], [21]. Here, we study the performance of PDRLS algorithm considering both sequential and stochastic partial-diffusion schemes under noisy information exchange.

A. Assumptions

Several simplifying assumptions have been traditionally adopted in the literature to gain insight into the performance

of such adaptive algorithms. To proceed with the analysis we shall therefore introduce similar assumptions to what has been used before in the literature, and use them to derive useful performance measures.

Assumptions. In order to make the analysis tractable, we introduce the following assumptions on statistical properties of the measurement data and noise signals.

- 1) The regression data $\mathbf{u}_{k,i}$ are temporally white and spatially independent random variables with zero mean and covariance matrix $\mathbf{R}_{u,k} \triangleq \mathbb{E}[\mathbf{u}_{k,i}^* \mathbf{u}_{k,i}] \geq 0$.
- 2) The noise signal $\mathbf{v}_k(i)$ and $\mathbf{v}_{k,i}^{(\psi)}$ are temporally white and spatially independent random variable with zero mean and co(variance) $\sigma_{v,k}^2$ and $R_{v,k}^{(\psi)}$, respectively. In addition, The quantities $\{R_{v,lk}^{(\psi)}\}$ are all zero if $l \in \mathcal{N}_k$ or when $l = k$.
- 3) The regression data $\{\mathbf{u}_{m,i_1}\}$, the model noise signals $\mathbf{v}_n(i_2)$, and the link noise signals $\mathbf{v}_{l_1 k_1, j_1}^{(\psi)}$ are mutually independent random variables for all indexes $\{i_1, i_2, j_1, k_1, l_1, m, n\}$.
- 4) In order to make the performance analysis tractable, we introduce the following ergodicity assumption. For sufficiently large i , at any node k , we can replace $\mathbf{P}_{k,i}$ and $\mathbf{P}_{k,i}^{-1}$ with their expected values, $\mathbb{E}[\mathbf{P}_{k,i}]$ and $\mathbb{E}[\mathbf{P}_{k,i}^{-1}]$, respectively.
- 5) For a sufficiently large i , at any node k , we have $\mathbb{E}[\mathbf{P}_{k,i}] = \mathbb{E}[\mathbf{P}_{k,i}^{-1}]^{-1}$.

Assumption 4 is a common assumption in the analysis of performance of RLS-type algorithm (see for example, [34], pp.318-319) and the results from assuming that the regressor vector of each node is an ergodic process, so that the time average of the nodes rank-one instantaneous regressor covariance matrix, $\mathbf{u}_{k,i}^* \mathbf{u}_{k,i}$, over a sufficiently long time range can be replaced with the ensemble average (expected value). Assumption 5 is a good approximation when λ is close to unity and the condition number of $\mathbf{R}_{u,k}$ is not very large [9], [15], [21], [34].

B. Network Update Equation

Our objective is to examine whether, and how fast, the weight estimates $\mathbf{w}_{k,i}$ from the distributed implementation (10), (11), and (20) converge towards the solution \mathbf{w}^o (5). To do so, we introduce the M1 error vectors:

$$\tilde{\boldsymbol{\psi}}_{k,i} \triangleq \mathbf{w}^o - \boldsymbol{\psi}_{k,i} \quad (24)$$

$$\tilde{\mathbf{w}}_{k,i} \triangleq \mathbf{w}^o - \mathbf{w}_{k,i} \quad (25)$$

Furthermore, denote the network intermediate estimate-error and estimate-error vector as

$$\tilde{\boldsymbol{\psi}}_i \triangleq \text{col} \{ \tilde{\boldsymbol{\psi}}_{1,i}, \dots, \tilde{\boldsymbol{\psi}}_{N,i} \} \quad (26)$$

$$\tilde{\mathbf{w}}_i \triangleq \text{col} \{ \tilde{\mathbf{w}}_{1,i}, \dots, \tilde{\mathbf{w}}_{N,i} \} \quad (27)$$

Also, collecting the noise signal (21) and its covariances from across the network into $N \times 1$ block vectors and $N \times N$ block

diagonal matrices as follows:

$$\mathbf{v}_i^{(\psi)} \triangleq \text{col} \{ \mathbf{v}_{1,i}^{(\psi)}, \dots, \mathbf{v}_{N,i}^{(\psi)} \} \quad (28)$$

$$\mathbf{R}_v^{(\psi)} \triangleq \text{col} \{ \mathbf{R}_{v,1}^{(\psi)}, \dots, \mathbf{R}_{v,N}^{(\psi)} \} \quad (29)$$

Using the data model (1) and subtracting \mathbf{w}^o from both sides of the relation (12) we obtain

$$\tilde{\boldsymbol{\psi}}_{k,i} = \tilde{\mathbf{w}}_{k,i-1} - \mathbf{P}_{k,i} \mathbf{u}_{k,i}^* [\mathbf{u}_{k,i} \tilde{\mathbf{w}}_{k,i-1} + \mathbf{v}_k(i)] \quad (30)$$

which can be written as

$$\begin{aligned} \tilde{\boldsymbol{\psi}}_{k,i} &= \mathbf{P}_{k,i} \mathbf{P}_{k,i}^{-1} \tilde{\mathbf{w}}_{k,i-1} - \mathbf{P}_{k,i} \mathbf{u}_{k,i}^* \mathbf{u}_{k,i} \tilde{\mathbf{w}}_{k,i-1} \\ &\quad - \mathbf{P}_{k,i} \mathbf{u}_{k,i}^* \mathbf{v}_k(i) \end{aligned} \quad (31)$$

rewriting (12) as

$$\mathbf{P}_{k,i}^{-1} = \lambda \mathbf{P}_{k,i-1}^{-1} + \mathbf{u}_{k,i}^* \mathbf{u}_{k,i} \quad (32)$$

Substituting (32) into the first term on RHS of (31) yields

$$\tilde{\boldsymbol{\psi}}_{k,i} = \lambda \mathbf{P}_{k,i} \mathbf{P}_{k,i-1}^{-1} \tilde{\mathbf{w}}_{k,i-1} - \mathbf{P}_{k,i} \mathbf{u}_{k,i}^* \mathbf{v}_k(i) \quad (33)$$

We are interested in the steady-state behavior of the matrix $\mathbf{P}_{k,i}$. As $i \rightarrow \infty$, and $0 \ll \lambda \leq 1$, the steady-state mean value of $\mathbf{P}_{k,i}^{-1}$ is given by

$$\begin{aligned} \lim_{i \rightarrow \infty} \mathbb{E}[\mathbf{P}_{k,i}] &= \lim_{i \rightarrow \infty} \mathbb{E} \left[\sum_{j=1}^i \lambda^{i-j} \mathbf{u}_{k,i}^* \mathbf{u}_{k,i} \right] \\ &= \lim_{i \rightarrow \infty} \sum_{j=1}^i \lambda^{i-j} \mathbb{E}[\mathbf{u}_{k,i}^* \mathbf{u}_{k,i}] \\ &= \left(\lim_{i \rightarrow \infty} \sum_{j=1}^i \lambda^{i-j} \right) \mathbf{R}_{u,k} \\ &= \frac{1}{1-\lambda} \mathbf{R}_{u,k} \end{aligned} \quad (34)$$

Using Assumptions (4) and (5) and relation (31), we have for large enough i

$$\begin{aligned} \mathbf{P}_{k,i} \mathbf{P}_{k,i-1}^{-1} &\approx \mathbb{E}[\mathbf{P}_{k,i}] \mathbb{E}[\mathbf{P}_{k,i-1}^{-1}] \\ &\approx \mathbb{E}[\mathbf{P}_{k,i}^{-1}]^{-1} \mathbb{E}[\mathbf{P}_{k,i-1}^{-1}] \\ &\approx \mathbf{I}_M \end{aligned} \quad (35)$$

and

$$\mathbf{P}_{k,i} \approx \mathbb{E}[\mathbf{P}_{k,i}] \approx \mathbb{E}[\mathbf{P}_{k,i}^{-1}]^{-1} \approx (1-\lambda) \mathbf{R}_{u,k}^{-1} \quad (36)$$

Consequently for a sufficiently large i , (33) can be approximated by

$$\tilde{\boldsymbol{\psi}}_{k,i} = \lambda \tilde{\mathbf{w}}_{k,i} - (1-\lambda) \mathbf{R}_{u,k}^{-1} \mathbf{u}_{k,i}^* \mathbf{v}_k(i) \quad (37)$$

On the other hand, subtracting both sides of (20) from \mathbf{w}^o gives

$$\begin{aligned} \tilde{\mathbf{w}}_{k,i} &= \left(\mathbf{I}_M - \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{lk} \mathbf{K}_{l,i} \right) \tilde{\boldsymbol{\psi}}_{k,i} \\ &\quad + \sum_{l \in \mathcal{N}_k \setminus \{k\}} a_{lk} \mathbf{K}_{l,i} \tilde{\boldsymbol{\psi}}_{l,i} - \mathbf{v}_{k,i}^{(\psi)} \end{aligned} \quad (38)$$

To describe these relations more compactly, we collect the information from across the network into block vectors and matrices. Using (26)-(28), leads to

$$\tilde{\psi}_i = \lambda \tilde{\mathbf{w}}_{i-1} - \mathbf{\Gamma} \mathbf{s}_i \quad (39)$$

$$\tilde{\mathbf{w}}_i = \mathbf{B}_i \tilde{\psi}_{i-1} - \mathbf{v}_i^{(\psi)} \quad (40)$$

where

$$\mathbf{\Gamma} \triangleq (1 - \lambda) \text{diag} \{ \mathbf{R}_{u,1}^{-1}, \dots, \mathbf{R}_{u,N}^{-1} \} \quad (41)$$

$$\mathbf{s}_i \triangleq \{ \mathbf{u}_{1,i}^* \mathbf{v}_1(i), \dots, \mathbf{u}_{N,i}^* \mathbf{v}_N(i) \} \quad (42)$$

$$\mathbf{B}_i = \begin{bmatrix} \mathbf{B}_{1,1,i} & \cdots & \mathbf{B}_{1,N,i} \\ \vdots & \ddots & \vdots \\ \mathbf{B}_{N,1,i} & \cdots & \mathbf{B}_{N,N,i} \end{bmatrix} \quad (43)$$

where

$$\mathbf{B}_{p,q,i} = \begin{cases} \mathbf{I}_M - \sum_{l \in \mathcal{N}_p \setminus \{p\}} a_{lp} \mathbf{K}_{l,i} & \text{if } p = q \\ a_{qp} \mathbf{K}_{q,i} & \text{if } q \in \mathcal{N}_p \setminus \{p\} \\ \mathbf{O}_M & \text{otherwise} \end{cases} \quad (44)$$

and \mathbf{O}_M is the $M \times M$ zero matrix. So, the network weight error vector, $\tilde{\mathbf{w}}_i$, ends up evolving according to the following stochastic recursion:

$$\tilde{\mathbf{w}}_i = \lambda \mathbf{B}_i \tilde{\mathbf{w}}_{i-1} - \mathbf{B}_i \mathbf{\Gamma} \mathbf{s}_i - \mathbf{v}_i^{(\psi)} \quad (45)$$

C. Convergence in Mean

Taking expectation of both sides of (45) under *Remark* and *Assumptions 1* and 2, we find that the mean error vector evolves according to the following recursion:

$$\mathbb{E}[\tilde{\mathbf{w}}]_i = \lambda \mathbf{Q} \mathbb{E}[\tilde{\mathbf{w}}_{i-1}] \quad (46)$$

where

$$\mathbf{Q} = \mathbb{E}[\mathbf{B}_i] \quad (47)$$

Like [20], \mathbf{Q} can be obtained for both stochastic and sequential partial-diffusion using the definition of \mathbf{B}_i . What is most noteworthy here is to find the value of each \mathbf{Q} entry after applying expectation operator. Therefore, we can write

$$\mathbb{E}[\mathbf{B}_{p,q,i}] = \begin{cases} (1 - \rho + \rho a_{r,lp}) \mathbf{I}_M & \text{if } p = q \\ \rho a_{r,qp} \mathbf{I}_M & \text{if } q \in \mathcal{N}_p \setminus \{p\} \\ \mathbf{O}_M & \text{otherwise} \end{cases} \quad (48)$$

All the entries of \mathbf{Q} are real and non-negative and all the rows of \mathbf{Q} add up to unity. This property can be established for both stochastic and sequential partial-diffusion schemes and for any value of L [20]. This implies that \mathbf{Q} is a right stochastic matrix. In view of the fact that the eigenvalue of a stochastic matrix with the largest absolute value (spectral radius) is equal to one [36], [41], the spectral radius of the matrix $\lambda \mathbf{Q}$ is equal to λ . Thus, for $0 \ll \lambda \leq 1$, every element of $\mathbb{E}[\tilde{\mathbf{w}}_i]$ converges to zero as $i \rightarrow \infty$, and the estimator is asymptotically unbiased and convergent in mean. Note that this is not in fact the necessary and sufficient condition for convergence of $\mathbb{E}[\tilde{\mathbf{w}}_i]$ as (46) has been obtained under an independence assumption which is not true in general.

D. Mean-Square Stability

We now study the mean-square performance of PDRLS under imperfect information exchange. We will also derive closed-form expressions that characterize the network performance. Doing so, we resort to the energy conservation analysis of [21], [35], [41], [42]. The details are as follows.

Equating the squared weighted Euclidean norm of both sides of (45) and applying the expectation operator together with using *Remark* and *Assumptions 1* and 2 yield the following weighted variance relation:

$$\mathbb{E}[\|\tilde{\mathbf{w}}_i\|_{\Sigma}^2] = \mathbb{E}[\tilde{\mathbf{w}}_{i-1}^* \lambda^2 \mathbf{B}_i^T \Sigma \mathbf{B}_i \tilde{\mathbf{w}}_{i-1}] + \mathbb{E}[\mathbf{s}_i^* \mathbf{\Gamma} \mathbf{B}_i^T \Sigma \mathbf{B}_i \mathbf{\Gamma} \mathbf{s}_i] + \mathbb{E}[\mathbf{v}_i^{*(\psi)} \Sigma \mathbf{v}_i^{(\psi)}] \quad (49)$$

where Σ is an arbitrary symmetric nonnegative-definite matrix. Let us evaluate each of the expectations on the right-hand side. The first expectation is given by

$$\begin{aligned} \mathbb{E}[\tilde{\mathbf{w}}_{i-1}^* \lambda^2 \mathbf{B}_i^T \Sigma \mathbf{B}_i \tilde{\mathbf{w}}_{i-1}] &= \mathbb{E}[\mathbb{E}[\tilde{\mathbf{w}}_{i-1}^* \lambda^2 \mathbf{B}_i^T \Sigma \mathbf{B}_i \tilde{\mathbf{w}}_{i-1} | \tilde{\mathbf{w}}_{i-1}]] \\ &= \mathbb{E}[\tilde{\mathbf{w}}_{i-1}^*] \mathbb{E}[\lambda^2 \mathbf{B}_i^T \Sigma \mathbf{B}_i] \\ &\triangleq \mathbb{E}[\tilde{\mathbf{w}}_{i-1}^* \Sigma' \tilde{\mathbf{w}}_{i-1}] \\ &= \mathbb{E}[\|\tilde{\mathbf{w}}_i\|_{\Sigma'}^2] \end{aligned} \quad (50)$$

where we introduce the nonnegative-definite weighting matrix

$$\Sigma' = \mathbb{E}[\lambda^2 \mathbf{B}_i^T \Sigma \mathbf{B}_i] \quad (51)$$

Since $\tilde{\mathbf{w}}_{i-1}$ is independent of Σ' , we have

$$\mathbb{E}[\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2] = \mathbb{E}[\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'}^2 | \Sigma'] \quad (52)$$

It is convenient to introduce the alternative notation $\|\mathbf{x}\|_{\sigma}^2$ to refer to the weighted square quantity $\|\mathbf{x}\|_{\Sigma}^2$, where $\sigma = \text{vec}\{\Sigma\}$. We shall use these two notations interchangeably.

Using the following equalities for arbitrary matrices $\{\mathbf{U}, \mathbf{W}, \Sigma, \mathbf{Z}\}$ of compatible dimensions:

$$(\mathbf{U} \otimes \mathbf{W})(\Sigma \otimes \mathbf{Z}) = \mathbf{U} \Sigma \otimes \mathbf{W} \mathbf{Z} \quad (53)$$

$$\text{vec}\{\mathbf{U} \Sigma \mathbf{W}\} = (\mathbf{W}^T \otimes \mathbf{U}) \text{vec}\{\Sigma\} \quad (54)$$

$$\text{Tr}(\Sigma \mathbf{W}) = [\text{vec}\{\mathbf{W}^T\}]^T \text{vec}\{\Sigma\} \quad (55)$$

we have

$$\sigma' = \mathbb{E}[\lambda^2 \mathbf{B}_i^T \Sigma \mathbf{B}_i] = \mathcal{F} \sigma \quad (56)$$

where

$$\mathcal{F} = \lambda^2 \Phi \quad (57)$$

$$\Phi = \mathbb{E}[\mathbf{B}_i^T \otimes \mathbf{B}_i^T] \quad (58)$$

Here, the efforts to find (58) in [20] can be extended. Therefore, Φ can be established in general form. Again, what is most important here is to discuss about the probability of transmitted entries of nodes. This would be helpful in finding some expectations, $\mathbb{E}[\kappa_{t,p,i} \mathbf{K}_{q,i}]$, we are faced with.

At a given iteration, the probability of transmitting two different entries of single node is [20]

$$\rho \left(\frac{L-1}{M-1} \right) \quad (59)$$

In the stochastic partial-diffusion scheme, at a given iteration, the probability of transmitting two entries from two different nodes is ρ^2 . Thus, we have [20]

$$\begin{aligned} \mathbb{E}[\kappa_{t,p,i} \mathbf{K}_{q,i}] &= \begin{cases} \rho \left(\frac{L-1}{M-1} \right) \mathbf{I}_M + \rho \left(\frac{M-L}{M-1} \right) \mathcal{J}_M^{t,t} & \text{if } p = q \\ \rho^2 \mathbf{I}_M & \text{if } p \neq q \end{cases} \end{aligned} \quad (60)$$

where $\mathcal{J}_M^{t,t}$ is an $M \times M$ single-entry matrix with one at its (t,t) th entry and zeros elsewhere, for $p = 1, \dots, N$, $q = 1, \dots, N$, and $t = 1, \dots, M$. Furthermore, in sequential partial-diffusion and under same entry selection pattern, at a given iteration, we have [20]

$$\mathbb{E}[\kappa_{t,p,i} \mathbf{K}_{q,i}] = \rho \left(\frac{L-1}{M-1} \right) \mathbf{I}_M + \rho \left(\frac{M-L}{M-1} \right) \mathcal{J}_M^{t,t} \quad (61)$$

As shown in [20], all the entries of Φ are real and non-negative. moreover, all their columns sum up to one both for stochastic and sequential partial-diffusion scheme and for any value of L .

Second term on RHS of (49)

$$\mathbb{E}[\mathbf{s}_i^* \mathbf{\Gamma} \mathbf{B}_i^T \Sigma \mathbf{B}_i \mathbf{\Gamma} \mathbf{s}_i] = \text{vec}^T \{ \mathcal{G} \} \Phi \sigma \quad (62)$$

where

$$\mathcal{G} = \mathbf{\Gamma} \mathbb{E}[\mathbf{s}_i \mathbf{s}_i^*] \mathbf{\Gamma} \quad (63)$$

which, in view of *Assumptions*, can be expressed as

$$\mathcal{G} = (1-\lambda)^2 \left\{ \sigma_{v,1}^2 \mathbf{R}_{u,1}^{-1}, \dots, \sigma_{v,N}^2 \mathbf{R}_{u,N}^{-1} \right\} \quad (64)$$

Last term on RHS of (49):

$$\begin{aligned} \mathbb{E}[\mathbf{v}_i^{*(\psi)} \Sigma \mathbf{v}_i^{(\psi)}] &= \mathbb{E}[\text{Tr} \{ \Sigma \mathbf{v}_i^{(\psi)} \mathbf{v}_i^{*(\psi)} \}] \\ &= \text{vec}^T \left(\mathbf{R}_v^{(\psi)} \right) \sigma \end{aligned} \quad (65)$$

The variance relation becomes

$$\begin{aligned} \mathbb{E}[\|\tilde{\mathbf{w}}_i\|_\sigma^2] &= \mathbb{E}[\|\tilde{\mathbf{w}}_{i-1}\|_{\lambda^2 \Phi \sigma}^2] + \text{vec}^T \{ \mathcal{G} \} \Phi \sigma \\ &\quad + \text{vec}^T \left(\mathbf{R}_v^{(\psi)} \right) \sigma \end{aligned} \quad (66)$$

A recursion of type (66) is stable and convergent if the matrix $\lambda^2 \Phi$ is stable [34]. The entries of Φ are all real-valued and non-negative. In addition, all the columns of Φ add up to unity. As a result, Φ is left-stochastic and has unit spectral radius. Hence, the spectral radius of the matrix $\lambda^2 \Phi$ is equal to λ^2 that is smaller than one. Therefore, the mean-square stability condition for PDRLS under noisy links is the same as that of PDRLS under noise-free links and the rate of this convergence is dependent on the value of λ . Again, this need not translate to be the necessary and sufficient condition for convergence of recursion of type (66) in actuality as (66) has been obtained under independence assumption which is not true in general. Therefore, the mean-square convergence of PDRLS under noisy links is an open question.

E. Mean-Square Performance

At steady state, when $i \rightarrow \infty$, (66) can be written as

$$\begin{aligned} \lim_{i \rightarrow \infty} \mathbb{E}[\|\tilde{\mathbf{w}}_i\|_{(\mathbf{I}_{N^2 M^2} - \mathcal{F})\sigma}^2] &= \\ &\text{vec}^T \{ \mathcal{G} \} \Phi \sigma + \text{vec}^T \left\{ \mathbf{R}_v^{(\psi)} \right\} \sigma \end{aligned} \quad (67)$$

Expression (66) is a very useful relation; it allows us to evaluate the network MSD through proper selection of the weighting vector σ . The network MSD is defined as the average value:

$$\text{MSD}^{\text{network}} \triangleq \lim_{i \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \mathbb{E}[\|\tilde{\mathbf{w}}_{k,i}\|^2] \quad (68)$$

which amounts to averaging the MSDs of the individual nodes. Therefore,

$$\text{MSD}^{\text{network}} = \lim_{i \rightarrow \infty} \frac{1}{N} \mathbb{E}[\|\tilde{\mathbf{w}}_i\|^2] = \lim_{i \rightarrow \infty} \mathbb{E}[\|\tilde{\mathbf{w}}_i\|_{1/N}^2] \quad (69)$$

This means that in order to recover the network MSD from relation (66), we should select the weighting vector σ such that

$$(\mathbf{I}_{N^2 M^2} - \mathcal{F}) \sigma = \frac{1}{N} \text{vec} \{ \mathbf{I}_{NM} \} \quad (70)$$

Solving for σ and substituting back into (66) we arrive at the following expression for the network MSD

$$\begin{aligned} \text{MSD}_{\text{noisy}}^{\text{network}} &= \frac{1}{N} \left[\text{vec}^T \{ \mathcal{G} \} \Phi + \text{vec}^T \left\{ \mathbf{R}_v^{(\psi)} \right\} \right] \\ &\quad (\mathbf{I}_{N^2 M^2} - \mathcal{F})^{-1} \text{vec} \{ \mathbf{I}_{NM} \} \end{aligned} \quad (71)$$

In perfect information exchange, the last two terms of (66) don not appear, so we can conclude that the network MSD deteriorates as follows:

$$\begin{aligned} \text{MSD}_{\text{noisy}}^{\text{network}} &= \text{MSD}_{\text{ideal}}^{\text{network}} + \\ &\quad \frac{1}{N} \text{vec}^T \left\{ \mathbf{R}_v^{(\psi)} \right\} (\mathbf{I}_{N^2 M^2} - \mathcal{F})^{-1} \text{vec} \{ \mathbf{I}_{NM} \} \end{aligned} \quad (72)$$

IV. SIMULATION RESULTS

In order to illustrate the performance of each PDRLS strategy under imperfect information exchange, we consider an adaptive network with a random topology and $N = 10$ where each node is, on average, connected to two other nodes. The measurements were generated according to model (1), and regressors, $\mathbf{u}_{k,i}$, were chosen Gaussian i.i.d with randomly generated different diagonal covariance matrices, $\mathbf{R}_{u,k}$. The additive noises at nodes are zero mean Gaussian with variances $\sigma_{v,k}^2$ and independent of the regression data. The traces of the covariance matrix regressors and the noise variances at all nodes, $\text{Tr}(\mathbf{R}_{u,k})$ and $\sigma_{v,k}^2$, are shown in Fig. 1. We also use white Gaussian link noise signals such that $\mathbf{R}_{v,lk}^{(\psi)} = \sigma_{\psi,lk}^2 \mathbf{I}_M$. All link noise variances $\sigma_{\psi,lk}^2 \mathbf{I}_M$, are randomly generated and illustrated in Fig. 2. We assign the link number by the following procedure. We denote the link from node l to node k as $\ell_{l,k}$, where $l \neq k$. Then, we collect the links $\{\ell_{l,k}, l \in \mathcal{N}_k \setminus \{k\}\}$ in an ascending order of l in the

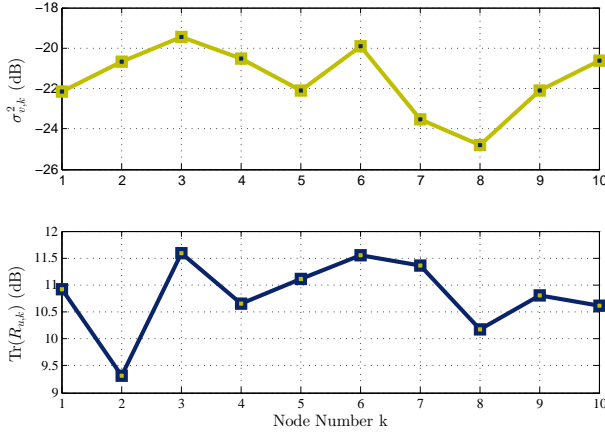


Fig. 1. Variance of the noise (top) and Covariance matrix trace of the input signal (bottom) at each node.

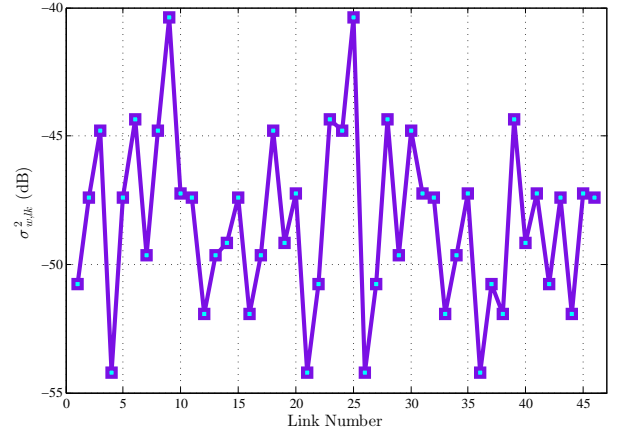


Fig. 2. The variance profiles for various sources of link noises in dB, $\sigma^2_{\psi, lk}$.

list \mathcal{L}_k (which is a set with ordered elements) for each node. We concatenate $\{\mathcal{L}_k\}$ in an ascending order of k to get the overall list $\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_N\}$. Eventually, the m th link in the network is given by the m th element in the list \mathcal{L} . It is noteworthy that we adopt the network MSD learning curves of all figures by averaging over 50 experiments and the unknown parameter w^o of length $M = 8$ is randomly generated. In Fig. 3, we illustrate the simulated time evolution of network MSD of the PDRLS algorithm under noisy information exchange using both sequential and stochastic partial-diffusion schemes for $\lambda = 0.995$. Fig. 4, demonstrates a similar scenario as that in Fig. 3 when $\lambda = 1$. To compare the steady-state performance of network MSD curves of PDRLS strategies, we examine the network MSD learning curve of PDRLS strategies again with ideal links under various numbers of entries communicated at each iteration, L , when $\lambda = 1$ as illustrated in Fig. 5. From the results above, we can make the following observations:

- The PDRLS algorithm delivers a tradeoff between communications cost and estimation performance under noise-free links [21];
- The MSD performance of PDRLS with noisy links is strictly come under the influence of forgetting factor, λ . So, a minimal change in λ leads to a significant degradation on MSD performance;
- As can be seen, PDLMS algorithm with noisy links fails to converge for both stochastic and sequential schemes, whereas it converges for the noise-free links case..
- The more entries are communicated at each iteration, the more perturbed weight estimates are interred in the consultation phase. Therefore, the number of communicated entries has a marked effect on the MSD performance of PDRLS with noisy links.

V. CONCLUSION AND FUTURE WORK

In this work, we investigated the performance of PDRLS algorithm when links between nodes were noisy. We derived an analytical expression for the network mean-square deviation, MSD, using weighted energy conservation relation. Our

results revealed that the noisy links are the main factor in the performance degradation of PDRLS algorithm. They also illustrated that the stability conditions for PDRLS under noisy links are not sufficient to guarantee its convergence. In other words, considering non-ideal links condition added a new complexity to the estimation problem for which the PDRLS algorithm became unstable and did not converge for any value of forgetting factor, λ . It was also showed that the PDLMS algorithm with noisy links shows divergent behavior for both selection schemes (stochastic and sequential), while it does not for the noise-free links case. In the future, tighter and accurate bounds on the convergence rate of the mean and mean-square update equations of PDRLS algorithm with noisy links can be established. Necessary and sufficient conditions for convergence of the algorithm with noisy links need to be derived. These can be addressed in the future.

REFERENCES

- [1] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *Signal Processing, IEEE Transactions on*, vol. 55, no. 8, pp. 4064–4077, 2007.
- [2] M. G. Rabbat and R. D. Nowak, "Quantized incremental algorithms for distributed optimization," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 4, pp. 798–808, 2005.
- [3] A. Nedic and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM Journal on Optimization*, vol. 12, no. 1, pp. 109–138, 2001.
- [4] A. Bertrand and M. Moonen, "Consensus-based distributed total least squares estimation in ad hoc wireless sensor networks," *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 2320–2330, 2011.
- [5] S. S. Stanković, M. S. Stanković, and D. M. Stipanović, "Decentralized parameter estimation by consensus based stochastic approximation," *Automatic Control, IEEE Transactions on*, vol. 56, no. 3, pp. 531–543, 2011.
- [6] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Distributed recursive least-squares for consensus-based in-network adaptive estimation," *Signal Processing, IEEE Transactions on*, vol. 57, no. 11, pp. 4583–4588, 2009.
- [7] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *Signal Processing, IEEE Transactions on*, vol. 60, no. 8, pp. 4289–4305, 2012.
- [8] X. Zhao, S.-Y. Tu, and A. H. Sayed, "Diffusion adaptation over networks under imperfect information exchange and non-stationary data," *Signal Processing, IEEE Transactions on*, vol. 60, no. 7, pp. 3460–3475, 2012.

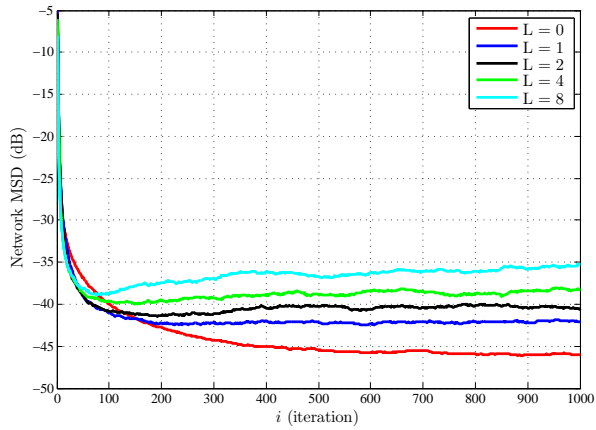
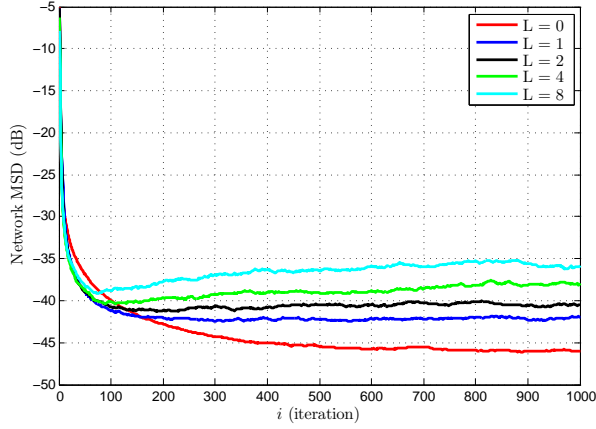


Fig. 3. Simulated network MSD curves for partial diffusion RLS algorithms using sequential (top) and stochastic (bottom) with different number of entries communicated under noisy links when $\lambda = 0.995$.

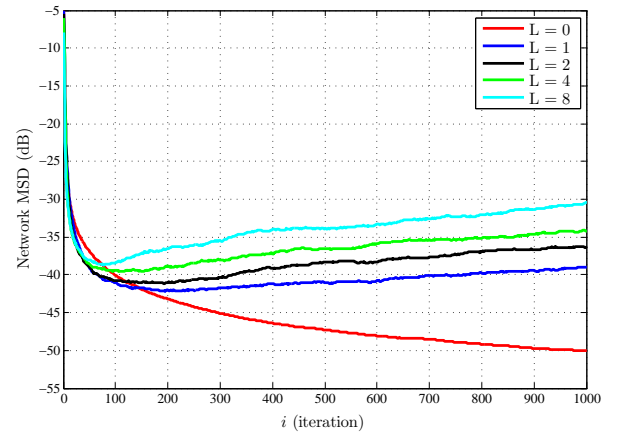
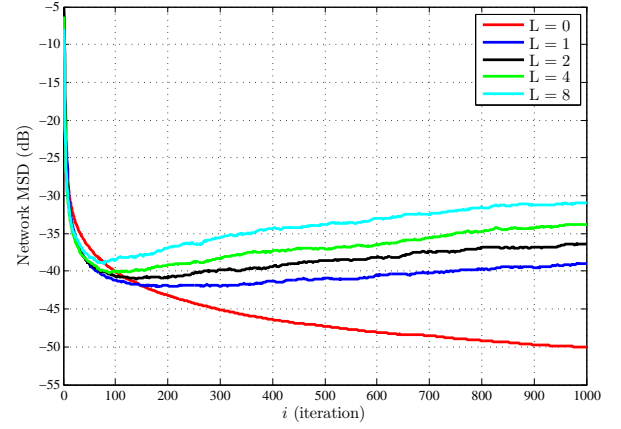


Fig. 4. Simulated network MSD curves for partial diffusion RLS algorithms using sequential (top) and stochastic (bottom) with different number of entries communicated under noisy links $\lambda = 1$.

- [9] A. Bertrand, M. Moonen, and A. H. Sayed, "Diffusion bias-compensated RLS estimation over adaptive networks," *Signal Processing, IEEE Transactions on*, vol. 59, no. 11, pp. 5212–5224, 2011.
- [10] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Adaptive robust distributed learning in diffusion sensor networks," *Signal Processing, IEEE Transactions on*, vol. 59, no. 10, pp. 4692–4707, 2011.
- [11] N. Takahashi, I. Yamada, and A. H. Sayed, "Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis," *Signal Processing, IEEE Transactions on*, vol. 58, no. 9, pp. 4795–4810, 2010.
- [12] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1035–1048, 2010.
- [13] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *Signal Processing, IEEE Transactions on*, vol. 56, no. 7, pp. 3122–3136, 2008.
- [14] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6217–6234, 2012.
- [15] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *Signal Processing, IEEE Transactions on*, vol. 56, no. 5, pp. 1865–1877, 2008.
- [16] Ø. L. Rørtveit, J. H. Husøy, and A. H. Sayed, "Diffusion LMS with communication constraints," in *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on*. IEEE, 2010, pp. 1645–1649.
- [17] C. G. Lopes and A. H. Sayed, "Diffusion adaptive networks with changing topologies," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008.
- [18] N. Takahashi and I. Yamada, "Link probability control for probabilistic diffusion least-mean squares over resource-constrained networks," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3518–3521.
- [19] X. Zhao and A. H. Sayed, "Single-link diffusion strategies over adaptive networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3749–3752.
- [20] R. Arablouei, S. Werner, Y.-F. Huang, and K. Dogancay, "Distributed least mean-square estimation with partial diffusion," *Signal Processing, IEEE Transactions on*, vol. 62, no. 2, pp. 472–484, 2014.
- [21] R. Arablouei, K. Dogancay, S. Werner, and Y.-F. Huang, "Adaptive distributed estimation based on recursive least-squares and partial diffusion," *Signal Processing, IEEE Transactions on*, vol. 62, no. 14, pp. 3510–3522, 2014.
- [22] M. O. Sayin and S. S. Kozat, "Single bit and reduced dimension diffusion strategies over distributed networks," *Signal Processing Letters, IEEE*, vol. 20, no. 10, pp. 976–979, 2013.
- [23] —, "Compressive diffusion strategies over distributed networks for reduced communication load," *Signal Processing, IEEE Transactions on*, vol. 62, no. 20, pp. 5308–5323, 2014.
- [24] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Trading off complexity with communication costs in distributed adaptive learning via Krylov subspaces for dimensionality reduction," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 7, no. 2, pp. 257–273, 2013.
- [25] R. Abdolee and B. Champagne, "Diffusion LMS algorithms for sensor networks over non-ideal inter-sensor wireless channels," in *Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on*. IEEE, 2011, pp. 1–6.
- [26] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chambers, "Tran-

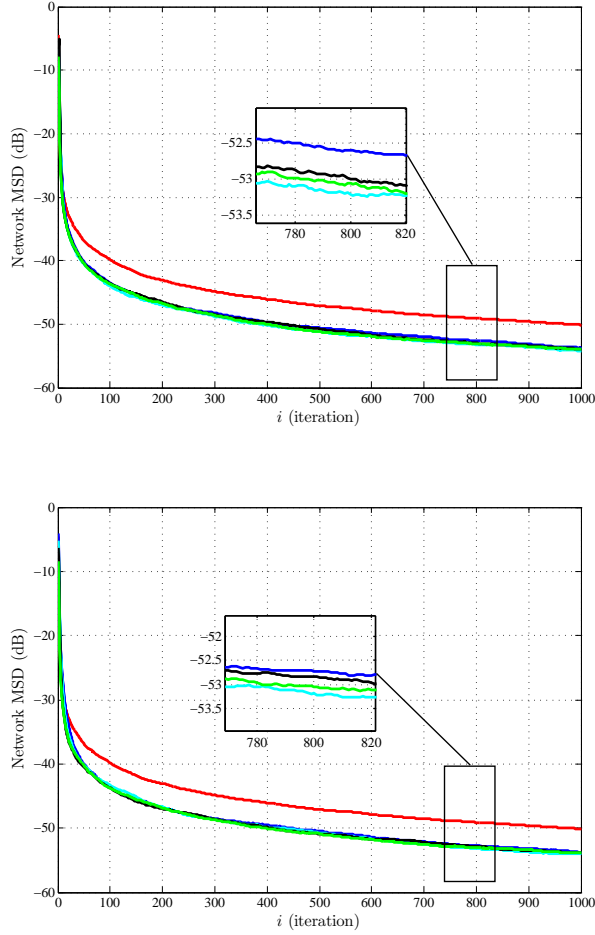


Fig. 5. Simulated network MSD curves for partial diffusion RLS algorithms using sequential (top) and stochastic (bottom) with different number of entries communicated under ideal links when $\lambda = 1$.

- [37] K. Dogancay, *Partial-update adaptive signal processing: Design Analysis and Implementation*. Academic Press, 2008.
- [38] M. Godavarti and A. O. Hero III, "Partial update LMS algorithms," *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, pp. 2382–2399, 2005.
- [39] S. C. Douglas, "Adaptive filters employing partial updates," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 44, no. 3, pp. 209–216, 1997.
- [40] J. R. Treichler, C. R. Johnson, and M. G. Larimore, *Theory and design of adaptive filters*. Wiley, 1987.
- [41] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of data-normalized adaptive filters," *Signal Processing, IEEE Transactions on*, vol. 51, no. 3, pp. 639–652, 2003.
- [42] N. R. Yousef and A. H. Sayed, "A unified approach to the steady-state and tracking analyses of adaptive filters," *Signal Processing, IEEE Transactions on*, vol. 49, no. 2, pp. 314–324, 2001.

sient analysis of diffusion leastmean squares adaptive networks with noisy channels," *International Journal of Adaptive Control and Signal Processing*, vol. 26, no. 2, pp. 171–180, 2012.

- [27] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. Chambers, "Steady-state analysis of diffusion LMS adaptive networks with noisy links," *Signal Processing, IEEE Transactions on*, vol. 60, no. 2, pp. 974–979, 2012.
- [28] S.-Y. Tu and A. H. Sayed, "Adaptive networks with noisy links," in *Global Telecommunications Conference (GLOBECOM 2011)*, 2011 IEEE. IEEE, 2011, pp. 1–5.
- [29] A. Khalili, M. A. Tinati, and A. Rastegarnia, "Performance analysis of distributed incremental LMS algorithm with noisy links," *International Journal of Distributed Sensor Networks*, vol. 2011, 2011.
- [30] —, "Steady-state analysis of incremental LMS adaptive networks with noisy links," *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 2416–2421, 2011.
- [31] —, "Analysis of incremental RLS adaptive networks with noisy links," *IEICE Electronics Express*, vol. 8, no. 9, pp. 623–628, 2011.
- [32] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *Signal Processing, IEEE Transactions on*, vol. 57, no. 1, pp. 355–369, 2009.
- [33] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Performance analysis of the consensus-based distributed LMS algorithm," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 68, 2009.
- [34] A. H. Sayed, *Adaptive filters*. John Wiley & Sons, 2011.
- [35] —, *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.
- [36] C. D. Meyer, *Matrix analysis and applied linear algebra*. Siam, 2000, vol. 2.